



CURATED STREAM

Harnessing AI in Data Analysis

Transforming insights into action

Agenda for today

What we'll cover

Six sections — from foundations through industry applications, ethics, and where the tools are heading.

1 Introduction to AI in data analysis

2 Core AI techniques

3 Applications in business

4 Case studies — healthcare, finance, retail

5 Challenges and ethical considerations

6 Future trends and 2026 tooling

SECTION 01

Introduction to AI in Data Analysis

Definitions, history, and why data is the lifeblood of every AI system.

What AI means in data analysis

Definition and scope

AI refers to the simulation of human-style intelligence by machines. In data analysis it means using algorithms — increasingly large language models — to analyse big datasets, identify patterns, and inform decisions without explicit, step-by-step programming.



Simulating intelligence

Algorithms and machine learning give machines the ability to perform cognitive tasks — classifying, predicting, reasoning over data.



Scaling data analysis

AI handles datasets too large or too messy for traditional spreadsheets, surfacing structure that humans would miss.



Pattern recognition

Models learn the patterns in your data and use them to predict, segment, or flag anomalies — often without being given the rules.

A short history of AI in analysis

From symbolic AI to today's frontier models

AI is not new — but its capability in data analysis has changed dramatically. Three waves drove it: symbolic rules, statistical machine learning, and today's foundation models.

1950s–1980s	1990s–2010s	2017–2022	2023–2026
Symbolic AI Hand-coded rules and expert systems. Brittle, but the first attempt to reason with computers.	Statistical ML Regression, decision trees, random forests, then deep neural networks. Powered by growing compute and big data.	Transformers Attention-based models unlock language and vision at scale. ChatGPT in late 2022 makes AI a household tool.	Frontier era Claude Fable 5 / Opus 4.8, GPT-5.6, Gemini 3.5 Flash — multimodal models analysing text, tables, images and video in one prompt.

Data is the lifeblood of AI

Why quality matters more than cleverness

Even the best model can only learn what its data shows it. Effective collection, cleaning and preprocessing are still the most under-rated steps in any AI project.



Data as the core component

Data underpins every step from training to deployment — model architecture matters less than the dataset behind it.



Quality drives performance

Accurate, relevant, representative data is what makes a model useful. "Garbage in, garbage out" still holds.



Collection and preprocessing

Cleaning, deduplication, labelling and feature design often consume 60–80% of a project's effort.

60–80%

of an AI project's effort goes into preparing data — not training models.

Source: industry rule of thumb (Forbes, IBM, Anaconda surveys)

SECTION 02

AI Techniques in Data Analysis

Machine learning, deep learning, and natural language processing — what each does well.

Machine learning algorithms

Learning from data to predict and classify

Machine learning powers most of the predictive analytics in business today. Algorithms learn from past data and improve as they see more.



Regression

Predict continuous outcomes like revenue, churn risk or temperature from a set of input variables.



Classification

Sort records into predefined classes — fraud / not fraud, customer segment, disease vs healthy.



Clustering

Group similar items together without labels — natural customer segments, anomalous transaction patterns.



Reinforcement

Agents learn by trial and error in an environment — used for pricing, logistics and game-playing AI.

Deep learning and neural networks

Where pattern complexity exceeds human design

Deep learning is a subset of machine learning that uses multi-layered neural networks. Each layer learns increasingly abstract features — letting the network handle problems that defeat classical methods, like understanding speech, images or natural text.



Neural network basics

Layers of simple units pass signals forward; training adjusts the connections to minimise error.



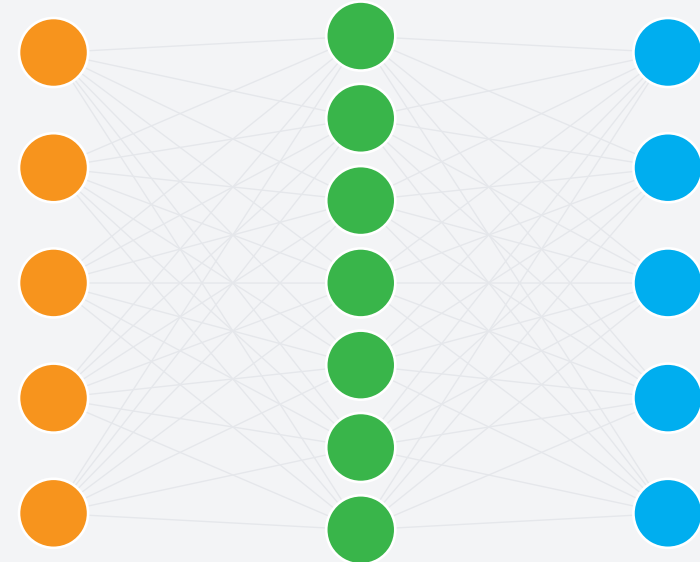
Vision and speech

Convolutional and transformer-based vision models now exceed humans on many image-recognition tasks.



Foundation models

Today's frontier models are deep networks at trillion-parameter scale, pre-trained on most of the public web.



Input → hidden layers → output

Each connection has a weight; training tunes weights to fit data.

Natural language processing

The bridge between humans and data

NLP lets machines understand and generate human language. With modern LLMs, NLP has become the dominant interface for analysing text, querying databases and summarising results.



Understanding language

Models parse meaning, intent and entities from free-form text — emails, tickets, reviews, contracts.



Text data analysis

Surface themes across millions of documents — what are customers really complaining about?



Sentiment analysis

Classify the emotional tone of feedback at scale — positive, neutral, negative, even mixed.



Conversational analytics

Ask a question in plain English; the model writes the SQL and explains the chart.

SECTION 03

Applications in Business

Where AI-driven data analysis actually creates value today.

Predictive analytics

Forecasting outcomes from historical data

Predictive analytics is the workhorse application of AI in business. It uses past data plus AI algorithms to forecast future events — sales, churn, demand, equipment failure, market risk.



Historical data

Patterns and trends in past data feed the forecast — quality and recency matter.



AI algorithms

Regression, time-series models and neural networks generate predictions over complex variables.



Better decisions

Anticipate market shifts and customer behaviour before they hit the bottom line.



Risk assessment

Flag potential downside risks early — credit, fraud, churn, supply-chain disruption.

Customer segmentation and personalisation

Targeted, not blanket

- AI clusters customers by behaviour and characteristics — far more nuanced than the classic demographic split.
- Marketing teams build campaigns that match each segment's needs, raising conversion and reducing wasted spend.
- Recommendation systems extend the same idea down to the individual — Netflix, Spotify, Amazon all rely on it.
- Modern LLMs turn segment definitions back into copy: describe the audience, get on-brand messaging in seconds.

Example segments

Loyal advocates

High lifetime value, high engagement

New explorers

Recent sign-ups still finding what they like

At-risk leavers

Engagement dropping over the past 30 days

Anomaly detection and fraud prevention

Finding the needle, not searching the haystack

AI is uniquely good at spotting the unusual in a sea of the routine. From card fraud to network intrusion to manufacturing defects, it lets organisations move from reactive to proactive risk management.



Why it matters

Anomalies often signal fraud, failure or attack — catching them early saves orders of magnitude in damage.



Pattern analysis

Models learn what "normal" looks like for each user, device or transaction stream — then flag the deviations.



Proactive mitigation

Automated alerts route to analysts; high-confidence cases can be blocked in real-time before they complete.

SECTION 04

Case Studies

Healthcare, financial services, retail — three industries reshaped by AI-driven analysis.

AI across three industries

Healthcare, financial services, retail

Different sectors apply the same techniques to different problems. The common thread: better predictions and better decisions at scale.



Healthcare

- Predictive modelling of patient outcomes
- Faster, more accurate diagnosis
- Personalised treatment recommendations
- Operational efficiency — wait times, triage

1,250

FDA-authorized AI medical devices by Jul 2025



Financial services

- Sharper risk assessment and credit scoring
- Algorithmic trading and market analysis
- Real-time fraud detection
- Portfolio optimisation

+90%

Rise in prompt-injection signals in customer deployments since late 2025



Retail and e-commerce

- Consumer behaviour analytics
- Demand forecasting and inventory optimisation
- Personalised recommendations
- Dynamic pricing

Up to NZ\$108B

Projected NZ GDP uplift from AI by 2038 (MBIE analysis)

SECTION 05

Challenges and Ethics

Privacy, bias, transparency — the questions to answer before you deploy.

Three things to get right before you deploy

Privacy, fairness, transparency



Data privacy and security

AI systems Hoover up personal data. Treat security as table stakes; an IBM 2025 report found 13% of organisations suffered AI-model or application breaches and 97% lacked proper AI access controls. The EU AI Act is fully applicable from 2 Aug 2026 — high-risk systems need conformity assessments, technical documentation and CE marking.



Bias and fairness

Bias creeps in through data selection, labelling choices and training objectives. Build evaluation sets that reflect your real users — including under-represented groups — and review outputs for discriminatory patterns before you ship.



Transparency and interpretability

AI models can act as black boxes. Be able to explain, at least roughly, why your model made a decision. Tools like SHAP, LIME and chain-of-thought traces help; in regulated settings they are required.

SECTION 06

Future Trends and 2026 Tooling

Where the field is heading — and what you can use today.

2026 tooling: analyse data with natural language

Two leading approaches, plus the long context that makes both work

By 2026 you no longer need to write Python to analyse a CSV. Two tool families dominate, each with a different sweet spot.



ChatGPT Code Interpreter / Codex

Upload CSV, Excel or JSON. The model writes and runs Python in-session to clean, analyse and chart your data. Still the leader for raw data crunching and reproducible workflows.



Claude Artifacts

Renders HTML / CSS / JS / React / SVG live in a side panel. Outstanding for interactive visualisations and dashboards. Claude Projects gives persistent context across long analyses.

Large context = bigger datasets in one prompt

1M

tokens — Claude Fable 5 / Opus 4.8 / Sonnet 5 default

1M

tokens — GPT-5.6 (up to 128K output)

10M

tokens — Llama 4 Scout (industry-leading)

Bringing it together

What to take away from this stream

Mastering AI in data analysis is about three things: knowing the techniques, choosing the right tool, and putting human judgement around the edges.



AI transforms data into action

From historical reporting to forward-looking prediction — the same data, far more value.



Tools empower organisations

Conversational analytics, Artifacts, Code Interpreter and 1M-token context windows put analysis within reach of every team.



Stay sharp on ethics and quality

Privacy, bias and transparency are not afterthoughts. The EU AI Act and your customers are watching.



The field keeps moving

What was state-of-the-art six months ago is now table stakes. Keep learning — these slides will be out of date soon.